



# Some multivariate risk indicators: minimization by using a Kiefer-Wolfowitz approach to the mirror stochastic algorithm

Peggy Cenac, Véronique Maume-Deschamps, Clémentine Prieur

## ► To cite this version:

Peggy Cenac, Véronique Maume-Deschamps, Clémentine Prieur. Some multivariate risk indicators: minimization by using a Kiefer-Wolfowitz approach to the mirror stochastic algorithm. *Statistics & Risk Modeling with Applications in Finance and Insurance*, 2012, 29 (1), pp.47-72. 10.1524/strm.2012.1069 . hal-00484233v2

**HAL Id: hal-00484233**

**<https://hal.science/hal-00484233v2>**

Submitted on 5 Dec 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SOME MULTIVARIATE RISK INDICATORS: MINIMIZATION BY USING A KIEFER-WOLFOWITZ APPROACH TO THE MIRROR STOCHASTIC ALGORITHM.

P. CÉNAC, V. MAUME-DESCHAMPS, AND C. PRIEUR

ABSTRACT. We consider some risk indicators of vectorial risk processes. These indicators take into account the dependencies between business lines as well as some temporal dependencies. By using stochastic algorithms, we may estimate the minimum of these risk indicators, under a fixed total capital constraint. This minimization may apply to capital reserve allocation.

## 1. INTRODUCTION

The new regulation rules for insurance industry, namely Solvency 2 in Europe, lead companies to adjust their solvency margins to the underlying risks. Once the overall company capital requirement has been computed, it must be splitted into solvency capitals for each line of business. In other words, given an initial capital  $u$  we assume that it is allocated to each line of business:  $u_k$  is the initial capital of the  $k^{\text{th}}$  line of business, then  $u_1 + \dots + u_d = u$ . We aim at optimizing the capital allocation with respect to some risk indicators. In this context  $u$  is fixed and we search for an optimal choice of the  $u_k$ 's. The ruin probability is a quite standard risk indicator and has been widely studied in dimension 1 (see [8] for a review). The ruin probability is related to the much used *Value at Risk* (VaR) which is simply a quantile and do not take into account the amount or severity of ruin, moreover it is not *coherent* in Artzner's *et al.* sense (see [1]). Risk measures incouting severity of ruin or some penalties functions have also been studied in dimension 1 (see for example [6, 23]) and they are related to the *Tail Value at Risk* (TVaR) or *Tail conditional Expectation* (see [1]). Generalizations of these notions in higher dimension are not straightforward. Several attempts exist. The simplest way to construct risk measures or risk indicators for vectors could be to consider risk measures or indicators associated to the sum of the coordinates (this is a way to aggregate risks). But, in a *ERM* (Enterprise Risk Management) point of view, one should wonder about the risk induced by each business line. Generalizations of the ruin probability or the tail conditional expectation may be found in [4, 10, 5].

---

*Key words and phrases.* Multivariate risk processes, risk indicators, stochastic algorithms, reserve allocation.

We would like to thank the referees for their careful reading of our manuscript and their comments. We are grateful to Anatoli Juditsky for his valuable advices. This work has been partially supported by the French Research National Agency (ANR) under the reference ANR-08-BLAN-0314-01.

In a multidimensional framework, risk indicators have been introduced by Loisel [26], which do not take into account the dependence structure or the ruin amount. We introduce some new risk indicators, both involving dependence and local insolvency severity. Considering that the main risk drivers for the overall company have been identified and that the global solvency capital requirement has been computed, they reveal the marginal solvency capitals for each line of business. A way to avoid as far as possible that some lines of business become insolvent too often could be to minimize these risk indicators, under a fixed total capital constraint. This might be achieved if some capital fungibility between lines of business or between entities is possible. One possible way to define optimality of the global reserve allocation is to minimize the expected sum of the penalties that each line of business would have to pay due to its temporary potential insolvency.

We consider a vectorial risk process  $X_i = (X_i^1, \dots, X_i^d)^t$ , where  $X_i^k$  corresponds to the gains of the  $k^{\text{th}}$  business line during the  $i^{\text{th}}$  period. That is,  $X_i^k = G_i^k - L_i^k$  where  $G_i^k$  denotes the incomes and  $L_i^k$  denotes the losses. We are interested in the cumulative gain:

$$Y_i^k = \sum_{p=1}^i X_p^k.$$

We shall say that the  $k^{\text{th}}$  line of business makes default during the  $n$  first periods if there exists  $i = 1, \dots, n$  such that  $Y_i^k + u_k < 0$ , where we recall that  $u_k$  is the initial capital of the  $k^{\text{th}}$  line of business. The ruin probability in a multivariate setting may be defined as the probability that one of the  $d$  lines of business makes default during the  $n$  first periods that is:

$$R(u_1, \dots, u_d) = \mathbb{P}(\exists k = 1, \dots, d \exists i = 1, \dots, n / Y_i^k + u_k < 0).$$

Other definitions of ruin probability and some bounds on them are provided in [4].

Let  $R_p^k$  denotes the capital of the  $k^{\text{th}}$  line of business at time  $p$ , that is  $R_p^k = u_k + Y_p^k$ . We may also be interested in the total cost of ruin :

$$A(u_1, \dots, u_d) = - \sum_{k=1}^d \mathbb{E} \left( \sum_{p=1}^n R_p^k \mathbf{1}_{\{R_p^k < 0\}} \right)$$

The asymptotic (as  $u \rightarrow \infty$ ) properties of  $A$  have been studied in [3]. More precisely, in the case  $d = 1$ , they study the asymptotics of  $A(u)$ . In higher dimension, they provide the asymptotic behavior of the vector  $u^* \in \mathbb{R}^d$  realizing the minimum of  $A$  under the constraint  $u_1 + \dots + u_d = u$ ,  $u_i \geq 0$ ,  $i = 1, \dots, d$ .

The risk indicator  $A$  does not take into account the dependence structure between the lines of business. A multivariate risk indicator that takes into account the dependence structure may be :

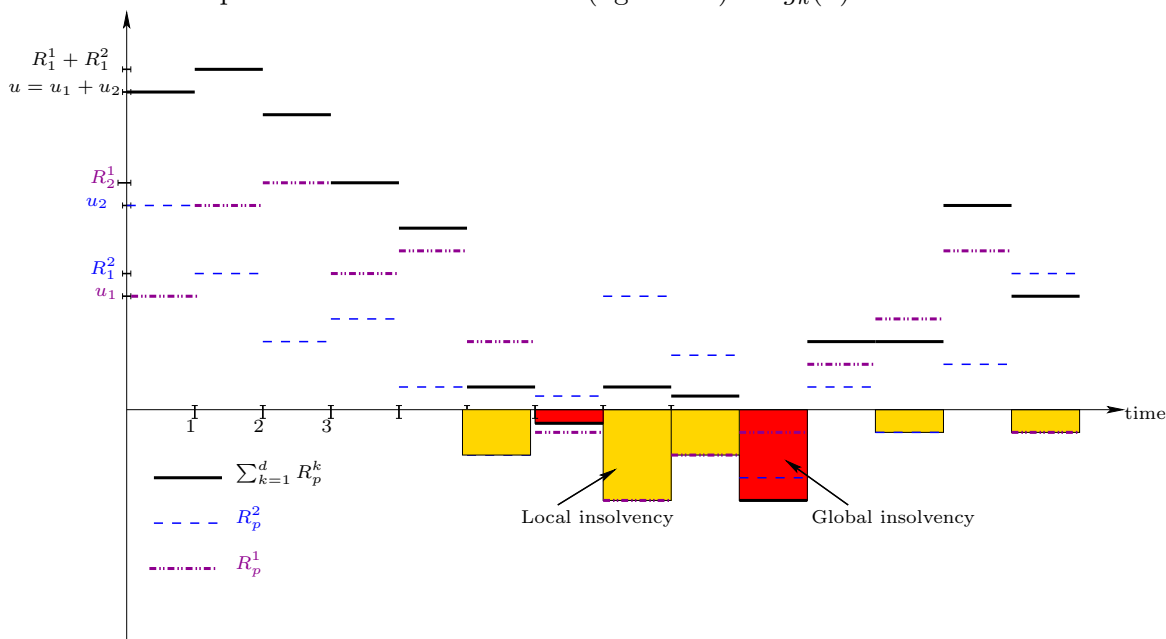
$$B(u_1, \dots, u_d) = \sum_{k=1}^d \mathbb{E} \left( \sum_{p=1}^n \mathbf{1}_{\{R_p^k < 0\}} \mathbf{1}_{\{\sum_{j=1}^d R_p^j > 0\}} \right).$$

This risk indicator is a measure of the time to ruin, it has been introduced in [26]. It does not take into account the severity (or the cost) of the insolvency. We introduce the following risk indicators, related to the ruin cost and depending on *penalty functions*  $g_k$ . Given convex functions  $g_k : \mathbb{R} \rightarrow \mathbb{R}$ ,  $k = 1, \dots, d$  satisfying  $g_k(x) \geq 0$  for  $x \leq 0$ , and  $\mathbb{E}(|g_k(R_p^k)|) < \infty$ , we consider the risk indicator

$$(1.1) \quad I(u_1, \dots, u_d) = \sum_{k=1}^d \mathbb{E} \left( \sum_{p=1}^n g_k(R_p^k) \mathbf{1}_{\{R_p^k < 0\}} \mathbf{1}_{\{\sum_{j=1}^d R_p^j > 0\}} \right).$$

The function  $g_k$  represents the cost that the  $k^{\text{th}}$  business branch has to pay when it becomes insolvent while the whole company is solvent. The risk indicator  $I$  is constructed with the idea that the solvency requirement for the whole company has been determined (namely  $u$  is determined). Minimizing  $I$  allows to control a partial insolvency, that is, insolvency measured for some branches while the whole company is solvent.

A natural choice for the cost could be  $g_k(x) = -x$ . The following picture represents the risk indicator  $I$  (light area) for  $g_k(x) = -x$  and  $d = 2$ .



We aim at finding a minimum under constraint on the initial capital  $u$ . Formally, we are looking for  $u^* \in (\mathbb{R}_+)^d$  such that

$$(1.2) \quad I(u^*) = \inf_{v_1 + \dots + v_d = u} I(v), \quad v \in (\mathbb{R}_+)^d.$$

Unless for very specific examples, no explicit solution is available. We propose to solve this problem by using stochastic algorithms (see [7, 2, 14] for a review). The classic algorithm introduced by Robbins and Monro [25] consists in applying a gradient descent method with a noisy gradient. In case such a noisy gradient is not observable, Kiefer and Wolfowitz (see [13]) developed a finite difference version of the Robbins-Monro method. The inconvenient of such algorithms is their very high sensitivity to the choice of the stepsize, which are very difficult to calibrate in practice. Moreover, we

are here in the context of constraint optimization problem. To handle the constraints, we first turned to the Lagrange multipliers approach. It was not tractable in practice in our setting, it did not converge on our simulations. Improvements of Robbins-Monro and Kiefer-Wolfowitz algorithms have been developed more recently using an averaging tool (see for instance [24, 22]) or with mirror versions of the stochastic descent algorithms. Deterministic form of the mirror descent algorithm is introduced by Nemirovski and Yudin [19]. The stochastic one has been proposed by Nesterov [21] and used in [11, 12] to solve non-smooth stochastic convex optimization problems. Recently new stochastic approximation algorithms, with stronger convergence properties, based on Nesterov's method have been proposed (see [9, 15, 16, 20]). All these algorithms rely on the existence of a *stochastic oracle*, that is a mechanism to generate noisy versions of the gradient. In our problem, we do not have access to a noisy gradient (or to a noisy subgradient). We are thus led to use approximations of  $\nabla I$  in stochastic algorithms. That is the reason why we have chosen to follow a Kiefer-Wolfowitz approach, in order to take into account an estimation of the gradient *online*. Our proof for the convergence of the algorithm is inspired by the work of Tauvel [27].

We construct a stochastic algorithm to estimate  $u^*$ , the *optimal allocation*. Under a moment condition (order  $> 2$ ) we prove the almost sure convergence of the estimator. Notice that our approach do not require any hypothesis, apart from a moment assumption, on the distribution of  $X_i$  and the existence of a density for some parts of our work. Also there may be temporal dependence on a period of length  $n$ . Nevertheless, in order to achieve the algorithm, we shall need  $N$  independent copies of the distribution of  $(X_1, \dots, X_n) \in (\mathbb{R}^d)^n$ , that is of the gains over a period of length  $n$ .

The paper is organized as follows: in section 2 we prove some convexity properties on  $I$ . In section 3 we make explicit the algorithm. In section 4 we prove the convergence results. In section 5 we show how to apply these results to our risk indicator and provide some simulations. While it is well known that standard Robbins-Monro and Kiefer-Wolfowitz algorithms are very sensitive to the choice of the stepsize, it is remarkable that our Kiefer-Wolfowitz version of the mirror algorithm is very stable. Namely, it seems, from our simulations, that any choice of the sequences in the theoretical convergence domain is convenient. So that our algorithm is very tractable while more standard ones are very hard to use in practice. For our purpose, more classical algorithms were not tractable, either because of the instability of the algorithm or because we do not have access to the gradient or even to an *oracle* of this gradient. Also, we have performed high dimensional simulations (10 and 40) and still the results are very satisfactory.

## 2. CONVEXITY PROPERTY

Let us first prove that, under suitable assumptions on  $g_k$ , the risk indicator  $I$  is convex.

$$\text{Let } v^{(-k)} = \sum_{j \neq k} v_j \text{ and } S_p = \sum_{k=1}^d Y_p^k.$$

**Proposition 2.1.** *Assume that the functions  $g_k$  are convex, that  $g_k : \mathbb{R} \rightarrow \mathbb{R}$  with  $g_k(0) = 0$ ,  $g_k(x) \geq 0$  for  $x < 0$ , and for all  $k = 1, \dots, d$  and  $p = 1, \dots, n$ ,  $g_k(R_p^k)$  admits an order 1 moment. Then the risk indicator  $I$  is convex on the convex set  $\mathcal{U}_u = \{(v_1, \dots, v_d) \in (\mathbb{R}^+)^d / v_1 + \dots + v_d = u\}$ . If moreover, for some  $k = 1, \dots, d$ ,  $g_k$  is strictly convex,  $g_k(x) > 0$  for  $x < 0$  and the joint distribution of  $(Y_p^k, S_p)$  contains  $] -u, +\infty[\times] -\infty, u[$  in its support then  $I$  admits a unique minimum in  $\mathcal{U}_u$ .*

*Proof.* We rewrite  $I$  as

$$I(u) = I(u_1, \dots, u_d) = \sum_{k=1}^d \mathbb{E} \left( \sum_{p=1}^n g_k(Y_p^k + u_k) \mathbf{1}_{\{Y_p^k < -u_k\}} \mathbf{1}_{\{S_p > -u\}} \right).$$

Since  $g_k(0) = 0$ ,

$$(u_1, \dots, u_d) \mapsto g_k(Y_p^k + u_k) \mathbf{1}_{\{Y_p^k < -u_k\}} \mathbf{1}_{\{S_p > -u\}}$$

is a convex function on  $\mathcal{U}_u$ . So that  $I$  is a convex function as the sum of expected values of convex functions. The unicity of the minimum follows from the fact that  $I$  is strictly convex, due to the strict convexity of the  $g_k$ 's.  $\square$

*Remark 1.* If, in addition to the hypothesis of Proposition 2.1, the functions  $g_k$  are differentiable and such that for all  $k = 1, \dots, d$  and  $p = 1, \dots, n$ ,  $g'_k(Y_p^k + u_k)$  admits an order 1 moment and for all  $i = 1, \dots, d$ ,  $(Y_p^i, S_p)$  has a joint density distribution denoted by  $f_{Y_p^i, S_p}$ , then the risk indicator  $I$  is differentiable. Indeed, it writes:

$$\begin{aligned} I(u) &= I(u_1, \dots, u_d) = \sum_{k=1}^d \mathbb{E} \left( \sum_{p=1}^n g_k(Y_p^k + u_k) \mathbf{1}_{\{Y_p^k < -u_k\}} \mathbf{1}_{\{S_p > -u\}} \right) \\ &= \sum_{k=1}^d \sum_{p=1}^n \int_{-\infty}^{-u_k} \int_{-u}^{\infty} g_k(y + u_k) f_{Y_p^k, S_p}(y, s) dy ds. \end{aligned}$$

Hence, the assumptions on  $g_k$  imply that  $I$  is differentiable. Moreover, we have:

$$\begin{aligned}
(\nabla I(v))_i &= \sum_{k=1}^d \sum_{p=1}^n \int_{-\infty}^{-v_k} g_k(y + v_k) f_{Y_p^k, S_p}(y, -u) dy \\
&\quad + \sum_{p=1}^n \mathbb{E}(g'_i(Y_p^i + v_i) \mathbf{1}_{\{Y_p^i < -v_i\}} \mathbf{1}_{\{S_p > -u\}}) \\
&\quad - \sum_{p=1}^n g_i(0) \int_{-u}^{+\infty} f_{Y_p^i, S_p}(-v_i, s) ds \\
&= \sum_{k=1}^d \sum_{p=1}^n \int_{-\infty}^{-v_k} g_k(y + v_k) f_{Y_p^k, S_p}(y, -u) dy \\
&\quad + \sum_{p=1}^n \mathbb{E}(g'_i(Y_p^i + v_i) \mathbf{1}_{\{Y_p^i < -v_i\}} \mathbf{1}_{\{S_p > -u\}}).
\end{aligned}$$

Thus  $\nabla I(v)$  is a sum of two terms. The second term appears as an expectation of an *observable* function which could be an oracle for a stochastic algorithm design. For the first part of the sum, such an oracle does not seem simply reachable, this is why we turned to a Kiefer-Wolfowitz approach.

The fact that  $I$  is a convex function is used to prove the convergence of our stochastic algorithm. The convexity assumption is also standard for classical stochastic algorithms (as Robbins-Monro or Kiefer-Wolfowitz) and also in more recent extensions ([9, 15, 16, 20]).

### 3. ESTIMATION BY USE OF STOCHASTIC ALGORITHMS

In this section we present our algorithm which is a Kiefer-Wolfowitz version of the mirror algorithm. In section 4 we shall prove its convergence under general assumptions (see Assumption 1) that are satisfied in particular for our specific risk indicator  $I$ .

**3.1. Auxiliary function for the mirror algorithm.** The stochastic mirror algorithm requires the use of an auxiliary function that will be used to push the trajectory into the set of constraints. We consider that  $\mathbb{R}^d$  is endowed with the  $L^1$  norm  $\|\cdot\|$  and the dual space  $(\mathbb{R}^d)^*$  is endowed with the dual norm  $\|\cdot\|_*$ :

$$\|x\| = \sum_{i=1}^d |x_i|, \quad \|\xi\|_* = \sup_{i=1, \dots, d} |\xi_i|.$$

Let  $C$  be a compact convex subset of  $\mathbb{R}^d$ . Recall that a convex function  $\delta : C \rightarrow \mathbb{R}$  is strongly convex with parameter  $\alpha > 0$  with respect to the norm  $\|\cdot\|$  if for any  $\lambda \in [0, 1]$  and any  $x, y$  belonging to  $C$ ,

$$\delta(\lambda x + (1 - \lambda)y) \leq \lambda \delta(x) + (1 - \lambda) \delta(y) - \frac{\alpha}{2} \lambda(1 - \lambda) \|x - y\|^2.$$

The mirror algorithm uses an  $\alpha$ -convex function  $\delta$  differentiable on a point  $x_0 \in C$  to define the auxiliary function  $V$ . Following Definition A.3 of [27], we consider:

$$V(x) = \delta(x) - \delta(x_0) - \langle \nabla \delta(x_0), x - x_0 \rangle.$$

It is then easy to see that the function  $V$  is also  $\alpha$ -convex.

Setting a positive  $\beta$ ,  $W_\beta$  denotes the Legendre-Fenchel transform of  $\beta V$  :

$$W_\beta(\xi) = \sup_{x \in C} \{ \langle \xi, x \rangle - \beta V(x) \}.$$

Due to the properties of Legendre-Fenchel transform of  $\alpha$ -convex functions, we know that  $W_\beta$  is continuously differentiable and that its gradient takes its values in  $C$ . Moreover, the gradient of  $W_\beta$  has a Lipschitz constant equal to  $(\alpha\beta)^{-1}$ .

We consider a function  $\Psi$  which is an approximation of the gradient of  $I$ . The mirror algorithm uses two positive sequences  $(\beta_n)_n$  and  $(\gamma_n)_n$  and a sequence of i.i.d. random vectors  $(\mathcal{Y}^n)_n$  in the following way (the choices for these sequences will be made explicit a bit later).

**Algorithm 1**

Initialization :	$\begin{cases} \xi_0 = 0 \in (\mathbb{R}^d)^\star \\ \chi_0 \in C \end{cases}$
Update :	<p>for <math>i = 1, \dots, N</math> do</p> $\begin{cases} \xi_i = \xi_{i-1} - \gamma_i \Psi(\chi_{i-1}, \mathcal{Y}_i) \\ \chi_i = \nabla W_{\beta_i}(\xi_i) \end{cases}$
Output :	$S_N = \frac{\sum_{i=1}^N \gamma_i \chi_{i-1}}{\sum_{i=1}^N \gamma_i}$

Under assumptions on the function  $\Psi$ , we shall prove that  $S^N$  goes a.s. to the unique minimum  $x^\star$  of  $I$ .

**3.2. Approximate gradient.** In our case, we do not have access directly to the gradient of  $I$ ,  $\nabla I$ , but we may approximate it *online*.

Recall that

$$I(u_1, \dots, u_d) = \sum_{k=1}^d \mathbb{E} \left( \sum_{p=1}^n g_k(R_p^k) \mathbf{1}_{\{R_p^k < 0\}} \mathbf{1}_{\{\sum_{k=1}^d R_p^k > 0\}} \right).$$

Let us recall that  $d$  denotes the number of lines of business, and  $n$  the number of periods of interest. The problem is to solve (1.2). We will consider an approximation  $\Psi$  of  $\nabla I$  with the following decomposition  $\Psi = \nabla I + \eta + r$ , where  $\eta$  is a martingale difference and  $r$  is negligible with respect to  $\nabla I$  and  $\eta$ . Arguing as for Kiefer-Wolfowitz, we first notice that

$$I(u_1, \dots, u_d) = \mathbb{E}(\mathcal{I}(u_1, \dots, u_d, \mathcal{Y}))$$



where

$$\mathcal{Y} = \begin{pmatrix} Y_1^1 & \cdots & Y_n^1 \\ \vdots & \ddots & \vdots \\ Y_1^d & \cdots & Y_n^d \end{pmatrix}$$

and

$$\mathcal{I}(u_1, \dots, u_d, y) = \sum_{k=1}^d \sum_{p=1}^n g_k(y_p^k + u_k) \mathbf{1}_{\{y_p^k + u_k < 0\}} \mathbf{1}_{\{\sum_{k=1}^d y_p^k + u_k > 0\}}.$$

Let us denote a realization of  $\mathcal{Y}$  by

$$y = \begin{pmatrix} y_1^1 & \cdots & y_n^1 \\ \vdots & \ddots & \vdots \\ y_1^d & \cdots & y_n^d \end{pmatrix}.$$

For sake of shortness, we introduce the following notations :

$$\begin{aligned} \mathcal{I}^k(c_i^+, \mathcal{Y}) &= \mathcal{I}(\chi_{i-1}^1, \dots, \chi_{i-1}^{k-1}, \chi_{i-1}^k + c_i, \chi_{i-1}^{k+1}, \dots, \chi_{i-1}^d, \mathcal{Y}), \\ \mathcal{I}^k(c_i^-, \mathcal{Y}) &= \mathcal{I}(\chi_{i-1}^1, \dots, \chi_{i-1}^{k-1}, \chi_{i-1}^k - c_i, \chi_{i-1}^{k+1}, \dots, \chi_{i-1}^d, \mathcal{Y}), \\ I^k(c_i^+) &= I(\chi_{i-1}^1, \dots, \chi_{i-1}^{k-1}, \chi_{i-1}^k + c_i, \chi_{i-1}^{k+1}, \dots, \chi_{i-1}^d), \\ I^k(c_i^-) &= I(\chi_{i-1}^1, \dots, \chi_{i-1}^{k-1}, \chi_{i-1}^k - c_i, \chi_{i-1}^{k+1}, \dots, \chi_{i-1}^d). \end{aligned}$$

We then consider  $D_{c_i} \mathcal{I}$  (*resp.*  $D_{c_i} I$ ) the random vector whose  $k^{\text{th}}$  coordinate  $D_{c_i}^k \mathcal{I}(u_1, \dots, u_d, \mathcal{Y})$  (*resp.*  $D_{c_i}^k I(u_1, \dots, u_d, \mathcal{Y})$ ) is defined by

$$\frac{\mathcal{I}^k(c_i^+, \mathcal{Y}) - \mathcal{I}^k(c_i^-, \mathcal{Y})}{2c_i} \quad \left( \text{resp. } \frac{I^k(c_i^+, \mathcal{Y}) - I^k(c_i^-, \mathcal{Y})}{2c_i} \right).$$

Let  $\Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i) = D_{c_i} \mathcal{I}(\chi_{i-1}, \mathcal{Y}_i)$ . Considering independent copies of the random matrix  $\mathcal{Y}$ , denoted by  $\mathcal{Y}_1, \dots, \mathcal{Y}_N$ , we perform Algorithm 2 below :

**Algorithm 2**

Initialization :	$\begin{cases} \xi_0 = 0 \in (\mathbb{R}^m)^* \\ \chi_0 \in C \end{cases}$
Update :	for $i = 1, \dots, N$ do  $\begin{cases} \xi_i = \xi_{i-1} - \gamma_i \Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i) \\ \chi_i = \nabla W_{\beta_i}(\xi_i) \end{cases}$
Output :	$S_N = \frac{\sum_{i=1}^N \gamma_i \chi_{i-1}}{\sum_{i=1}^N \gamma_i}$

Then defining

$$\begin{aligned} \eta_{c_i}(\chi_{i-1}, \mathcal{Y}_i) &= D_{c_i} \mathcal{I}(\chi_{i-1}, \mathcal{Y}_i) - D_{c_i} I(\chi_{i-1}), \\ r_{c_i}(\chi_{i-1}) &= D_{c_i} I(\chi_{i-1}) - \nabla I(\chi_{i-1}), \end{aligned}$$

it enhances the decomposition

$$(3.1) \quad \Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i) = \nabla I(\chi_{i-1}) + \eta_{c_i}(\chi_{i-1}, \mathcal{Y}_i) + r_{c_i}(\chi_{i-1}).$$

In the next section, we give conditions on  $(\gamma_i)_i$ ,  $(\beta_i)_i$  and  $(c_i)_i$  for the convergence of the algorithm.

#### 4. CONVERGENCE OF THE ALGORITHM

Let us consider  $C$  a compact convex subset of  $\mathbb{R}^d$ . We make also the following assumptions on  $I$  :

**Assumption 1.**

- (1)  $I$  is a convex function from  $\mathbb{R}^d$  to  $\mathbb{R}$ ,
- (2)  $I$  is of  $C^2$  class,
- (3)  $I$  admits a unique minimum  $x^*$  on  $C$ ,

Moreover, we have for any  $x$  in  $\mathbb{R}^d$  and any  $c > 0$

$$\mathbb{E}(\eta_c(x, \mathcal{Y})) = 0 \quad \text{and}$$

$r_c(x)$  converges to zero as  $c$  goes to 0.

We perform **Algorithm 2** with  $\chi_0$  being any random vector taking values in  $C$ , and  $\mathcal{F}_i$  being the  $\sigma$ -field generated by  $\chi_0, \mathcal{Y}_1, \dots, \mathcal{Y}_i$ . Then  $\xi_i$  is  $\mathcal{F}_i$  measurable and  $\mathcal{Y}_i$  is independent of  $\mathcal{F}_{i-1}$ . We get therefore

$$\mathbb{E}(\eta_{c_i}(\chi_{i-1}, \mathcal{Y}_i) | \mathcal{F}_{i-1}) = 0.$$

We shall need the following additional assumption on the function  $\mathcal{I}$ .

**Assumption 2.** There exists some nonnegative real number  $\sigma$  such that for all vector  $(v_1, \dots, v_d) \in \mathbb{R}^d$ , the random variable  $\mathcal{I}(x, \mathcal{Y})$  satisfies the moment condition  $\forall i$

$$\text{Var}(\mathcal{I}(v_1, \dots, v_d, \mathcal{Y}_i) | \mathcal{F}_{i-1}) \leq \sigma^2.$$

*Remark 2.* Assumption 2 yields the existence of a real number  $\sigma > 0$  such that for any  $i = 1, \dots, N$ ,

$$(4.1) \quad \mathbb{E} \left( \|\eta_{c_i}(\chi_{i-1}, \mathcal{Y}_i)\|_*^2 | \mathcal{F}_{i-1} \right) \leq d \frac{\sigma^2}{c_i^2}.$$

Indeed, one has

$$\begin{aligned} \mathbb{E} \left( \|\eta_{c_i}(\chi_{i-1}, \mathcal{Y}_i)\|_*^2 | \mathcal{F}_{i-1} \right) &\leq \sum_{k=1}^d \left( \frac{\text{Var}(\mathcal{I}^k(c_i^+, \mathcal{Y}_i) | \mathcal{F}_{i-1})}{4c_i^2} + \frac{\text{Var}(\mathcal{I}^k(c_i^-, \mathcal{Y}_i) | \mathcal{F}_{i-1})}{4c_i^2} \right. \\ &\quad \left. - \frac{1}{2c_i^2} \mathbb{E} \left[ \left( \mathcal{I}^k(c_i^+, \mathcal{Y}_i) - I(c_i^+) \right) \left( \mathcal{I}^k(c_i^-, \mathcal{Y}_i) - I(c_i^-) \right) \right] \right). \end{aligned}$$

From Assumption 2 together with the Cauchy-Schwartz inequality, one gets

$$\mathbb{E} \left( \|\eta_{c_i}(\chi_{i-1}, \mathcal{Y}_i)\|_*^2 | \mathcal{F}_{i-1} \right) \leq d \frac{\sigma^2}{c_i^2},$$

which concludes the proof of (4.1). Condition (4.1) is sufficient to obtain the  $L^1$ -convergence. In order to prove the almost sure convergence, we shall need a higher order moment condition.

**Assumption 3.** There exists  $p > 2$  such that almost surely

$$\sup_{i \geq 0} \mathbb{E}(|\mathcal{I}(v_1, \dots, v_d, \mathcal{Y})|^p | \mathcal{F}_{i-1}) < \infty.$$

*Remark 3.* Assumption 3 yields the existence of a real  $p > 2$  such that almost surely

$$\sup_{i \geq 0} \mathbb{E}(\|c_i \eta_{c_i}(\chi_{i-1}, \mathcal{Y}_i)\|_{\star}^p | \mathcal{F}_{i-1}) < \infty.$$

Our main result is the following.

**Theorem 4.1.** *Let  $(\beta_i)_{i \geq 0}$ ,  $(\gamma_i)_{i \geq 0}$  and  $(c_i)_{i \geq 0}$  be sequences in  $(\mathbb{R}_+^*)^{\mathbb{N}}$ . Assume moreover that  $(\beta_i)_{i \geq 0}$  is non decreasing and that the following assumptions are satisfied :*

- (i)  $\beta_N / \sum_{i=1}^N \gamma_i \xrightarrow{N \rightarrow +\infty} 0$ ,
- (ii)  $\sum_{i=1}^N \gamma_i c_i / \sum_{i=1}^N \gamma_i \xrightarrow{N \rightarrow +\infty} 0$ ,
- (iii)  $\sum_{i=1}^N \frac{\gamma_i^2}{c_i^2 \beta_{i-1}} / \sum_{i=1}^N \gamma_i \xrightarrow{N \rightarrow +\infty} 0$ ,
- (iv)  $\sum_{i=1}^{+\infty} \left( \frac{\gamma_i}{c_i} \right)^2 < \infty$ .

Then, provided that Assumptions 1 and 2 are satisfied, we have :

$$S_N \xrightarrow{L^1} x^*.$$

If moreover, Assumption 3 is satisfied, we have :

$$S_N \xrightarrow{a.s.} x^*.$$

*Remark 4.* Let  $0 < c < a - \frac{1}{2} < \frac{1}{2}$ . One can choose for  $(\beta_i)_{i \geq 0}$  the constant sequence equal to 1, for  $(c_i)_{i \geq 0}$  the sequence  $(i^{-c})_{i \geq 0}$  and for  $(\gamma_i)_{i \geq 0}$  the sequence  $(i^{-a})_{i \geq 0}$ .

*Remark 5.* For the “classical” Kiefer-Wolfowitz algorithm, the assumptions on the gains are particular cases of our assumption with a constant sequence  $(\beta_i)_{i \geq 0}$ :

$$\sum \gamma_n = \infty, \quad \sum \gamma_n c_n < \infty, \quad \text{and} \quad \sum \gamma_n^2 c_n^{-2} < \infty.$$

The proof of Theorem 4.1 uses an estimation of  $\varepsilon_N = I(S_N) - I(x^*)$  inspired from [27].

**Lemma 4.2.** *Assume that  $(\beta_i)_{i \in \mathbb{N}}$  is a non decreasing sequence, then*

$$\begin{aligned} 0 \leq \left( \sum_{i=1}^N \gamma_i \right) \varepsilon_N &\leq \beta_N V(x^*) - \sum_{i=1}^N \gamma_i \langle \eta_{c_i}(\chi_{i-1}, \mathcal{Y}_i), \chi_{i-1} - x^* \rangle \\ &\quad - \sum_{i=1}^N \gamma_i \langle r_{c_i}(\chi_{i-1}), \chi_{i-1} - x^* \rangle + \sum_{i=1}^N \frac{\gamma_i^2}{2\alpha\beta_{i-1}} \|\Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i)\|_{\star}^2. \end{aligned}$$

*Proof.* We remark that because  $I(x^*) = \inf_{x \in C} I(x)$  then  $\varepsilon_N \geq 0$  and for any  $i = 1, \dots, N$ ,  $I(\chi_{i-1}) - I(x^*) \geq 0$ . Because  $I$  is a convex and differentiable

function,

$$\begin{aligned} \left( \sum_{i=1}^N \gamma_i \right) \varepsilon_N &\leq \sum_{i=1}^N \gamma_i (I(\chi_{i-1}) - I(x^*)) \\ &\leq \sum_{i=1}^N \gamma_i \langle \nabla I(\chi_{i-1}), \chi_{i-1} - x^* \rangle. \end{aligned}$$

Thus,

$$\begin{aligned} \left( \sum_{i=1}^N \gamma_i \right) \varepsilon_N &\leq \sum_{i=1}^N \gamma_i \langle \Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i), \chi_{i-1} - x^* \rangle \\ &\quad - \sum_{i=1}^N \gamma_i \langle \eta_{c_i}(\chi_{i-1}, \mathcal{Y}_i), \chi_{i-1} - x^* \rangle - \sum_{i=1}^N \gamma_i \langle r_{c_i}(\chi_{i-1}), \chi_{i-1} - x^* \rangle \\ &= \langle \xi_N, x^* \rangle + \sum_{i=1}^N \gamma_i \langle \Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i), \chi_{i-1} \rangle \\ &\quad - \sum_{i=1}^N \gamma_i \langle \eta_{c_i}(\chi_{i-1}, \mathcal{Y}_i), \chi_{i-1} - x^* \rangle - \sum_{i=1}^N \gamma_i \langle r_{c_i}(\chi_{i-1}), \chi_{i-1} - x^* \rangle \end{aligned}$$

Now, because the sequence  $(\beta_i)_{i \in \mathbb{N}}$  is non decreasing and using the convexity and the differentiability of  $W_{\beta_{i-1}}$  (as the Legendre transform of a strictly convex function), Lemma A.1. of Tauvel [27] yields :

$$\begin{aligned} W_{\beta_i}(\xi_i) - W_{\beta_{i-1}}(\xi_{i-1}) &\leq W_{\beta_{i-1}}(\xi_i) - W_{\beta_{i-1}}(\xi_{i-1}) \\ &\leq \langle \nabla W_{\beta_{i-1}}(\xi_{i-1}), \xi_i - \xi_{i-1} \rangle + \frac{\gamma_i^2}{2\alpha\beta_{i-1}} \|\Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i)\|_*^2 \\ &= -\gamma_i \langle \chi_{i-1}, \Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i) \rangle + \frac{\gamma_i^2}{2\alpha\beta_{i-1}} \|\Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i)\|_*^2. \end{aligned}$$

As a consequence, we get:

$$\begin{aligned} \sum_{i=1}^N \gamma_i \langle \Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i), \chi_{i-1} \rangle &\leq W_{\beta_0}(\xi_0) - W_{\beta_N}(\xi_N) \\ &\quad + \sum_{i=1}^N \frac{\gamma_i^2}{2\alpha\beta_{i-1}} \|\Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i)\|_*^2 \\ &= -W_{\beta_N}(\xi_N) + \sum_{i=1}^N \frac{\gamma_i^2}{2\alpha\beta_{i-1}} \|\Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i)\|_*^2. \end{aligned}$$

The result finally follows from:

$$\langle x^*, \xi_N \rangle - W_{\beta_N}(\xi_N) \leq \beta_N V(x^*).$$

□

*Remark 6.* With our assumptions, we have that

$$\mathbb{E}(\langle \eta_{c_i}(\chi_{i-1}, \mathcal{Y}_i), \chi_{i-1} - x^* \rangle) = 0,$$

indeed,  $(\langle \eta_{c_i}(\chi_{i-1}, \mathcal{Y}_i), \chi_{i-1} - x^* \rangle)_i$  is a martingale difference sequence.

The following lemma may be deduced from Taylor inequality, recalling that  $\chi_i \in C \forall i$ .

**Lemma 4.3.** *There exists  $\kappa > 0$  such that, for any  $i$ ,*

$$\|r_{c_i}(\chi_{i-1})\|_* \leq \kappa c_i.$$

*Proof.* It follows from the fact that the Hessian matrix of  $I$  on  $C$  is bounded.  $\square$

*Remark 7.* As a consequence of Lemmas 4.2 and 4.3, we have that (recall that  $\varepsilon_N \geq 0$ )

$$0 \leq \mathbb{E}(|\varepsilon_N|) = \mathbb{E}(\varepsilon_N) \leq \frac{\beta_N V(x^*)}{\sum_{i=1}^N \gamma_i} + \frac{2\kappa D \sum_{i=1}^N \gamma_i c_i}{\sum_{i=1}^N \gamma_i} + \frac{\sum_{i=1}^N \frac{B\gamma_i^2}{2c_i^2 \alpha \beta_{i-1}}}{\sum_{i=1}^N \gamma_i},$$

where  $D$  is the diameter of  $C$  and  $\frac{B}{c_i^2}$  denotes a bound of  $\mathbb{E} \|\Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i)\|_*^2$ . Indeed, because of Decomposition (3.1)

$$(4.2) \quad \begin{aligned} \|\Psi_{c_i}\|_*^2 &\leq \|\nabla I\|_*^2 + \|r_{c_i}\|_*^2 + \|\eta_{c_i}\|_*^2 \\ &\quad + 2((\|\nabla I\|_* + \|r_{c_i}\|_*) \|\eta_{c_i}\|_* + \|\nabla I\|_* \cdot \|r_{c_i}\|_*). \end{aligned}$$

Recall that  $\|\nabla I\|_*$  is bounded (because  $C$  is compact and  $\nabla I$  is continuous), also following Lemma 4.3,

$$\|r_{c_i}\|_* \leq \kappa c_i.$$

Finally, using (4.1),

$$\mathbb{E}(\|\eta_{c_i}\|_*^2) \leq \frac{d\sigma^2}{c_i^2}.$$

So that, sufficient conditions to have the  $L^1$ -convergence of  $(\varepsilon_N)$  are Assumptions (i), (ii) and (iii) of Theorem 4.1.

We shall apply some standard results for martingales. Let us fix some notations :

**Definition 1.** Let  $M_N$  be a square integrable martingale written as

$$M_N = \sum_{i=1}^N e_i$$

with  $(e_i)_i$  a martingale difference sequence. Let

$$\langle M \rangle_N = \sum_{i=1}^N \mathbb{E}(e_i^2 | \mathcal{F}_{i-1}),$$

with  $\mathcal{F}_{i-1} = \sigma(e_1, \dots, e_{i-1})$ .

The following results are well known.

**Theorem** (distribution of large numbers for martingales, see for instance [18]). *Let  $M_N$  be a square integrable martingale if*

$$\langle M \rangle_\infty = \lim_{N \rightarrow \infty} \langle M_N \rangle < \infty \text{ a.s.}$$

*then  $(M_N)_{N \in \mathbb{N}}$  converges a.s. to a square integrable  $M_\infty$ .*

**Theorem** (Chow Lemma, see for instance [7] p.22). *Suppose  $(a_N)_{N \in \mathbb{N}}$  is a bounded sequence of positive random variables, suppose that  $1 < p \leq 2$ . For  $N \in \mathbb{N}$ , let  $A_N = 1 + \sum_{k=0}^N a_k$  and  $A_\infty = \lim_{N \rightarrow \infty} A_N$ . Suppose that  $(Z_N)_{N \in \mathbb{N}}$  is a positive sequence of random variables adapted to  $\mathcal{F}_N$  and  $K$  is a constant such that*

$$\mathbb{E}(Z_{N+1}|\mathcal{F}_N) \leq K \text{ and } \sup_N \mathbb{E}(Z_{N+1}^p|\mathcal{F}_N) < \infty$$

*then we have the following properties almost surely :*

$$(4.3) \quad \text{on } \{A_\infty < \infty\} \quad \sum_{k=1}^{\infty} a_k Z_{k+1} \text{ converges}$$

$$(4.4) \quad \text{on } \{A_\infty = \infty\} \quad A_N^{-1} \sum_{k=1}^N a_k Z_{k+1} \leq K.$$

*Proof of Theorem 4.1.* We have proved the convergence in  $L^1$  of  $(\varepsilon^N)_N$  to 0 (see Remark 7). The convergence of  $S_N$  to  $x^*$  follows:  $I(S_N)$  converges in  $L^1$  to  $I(x^*)$ , so it converges also in probability to  $I(x^*)$ . The unicity of  $x^*$  (as a minimum of  $I$ ) and the continuity of  $I$  gives that for any  $\delta > 0$ , there exists  $\epsilon > 0$  such that:

$$|s - x^*| > \delta \implies |I(s) - I(x^*)| > \epsilon.$$

Now, since  $S_N$  and  $x^*$  belong to the compact set  $C$ , there exists  $K > 0$  such that for any  $N \in \mathbb{N}$  and any  $\delta$ ,

$$\begin{aligned} \mathbb{E}|S_N - x^*| &\leq \delta + |S_N - x^*| \mathbb{P}(|S_N - x^*| > \delta) \\ &\leq \delta + K \mathbb{P}(|S_N - x^*| > \delta) \\ &\leq \delta + K \mathbb{P}(|I(S_N) - I(x^*)| > \epsilon) \end{aligned}$$

Thus, for any  $\delta$ ,  $\limsup_{N \rightarrow \infty} \mathbb{E}|S_N - x^*| \leq \delta$  (using the convergence in probability)

and finally we get the convergence in  $L^1$  of  $S_N$  to  $x^*$ .

Let us prove the almost sure convergence. We apply the distribution of large numbers for martingales to

$$M_N = \sum_{i=1}^N \gamma_i < \eta_{c_i}(\chi_{i-1}, \mathcal{Y}_i), \chi_{i-1} - x^* >,$$

$M_N$  is a martingale. Under Assumption (iv),  $\langle M_N \rangle$  is bounded and hence converges because it is increasing, thus  $M_N$  converges almost surely. Using

Lemma 4.2 we get:

$$\begin{aligned}
 (4.5) \quad 0 \leq \varepsilon_N &\leq \frac{\beta_N V(x^*)}{\sum_{i=1}^N \gamma_i} - \frac{M_N}{\sum_{i=1}^N \gamma_i} + \frac{2D\kappa \sum_{i=1}^N c_i \gamma_i}{\sum_{i=1}^N \gamma_i} \\
 (4.6) \quad &+ \frac{\sum_{i=1}^N \frac{\gamma_i^2}{2\alpha\beta_{i-1}} \|\Psi_{c_i}(\chi_{i-1}, \mathcal{Y}_i)\|_\star^2}{\sum_{i=1}^N \gamma_i}.
 \end{aligned}$$

The three terms in (4.5) go to zero a.s. by using the convergence of  $M_N$ , (i) and (ii), remark that (i) implies that  $\sum \gamma_i = +\infty$ . In order to prove that the term in (4.6) converges to zero a.s., we use the decomposition (4.2) and we apply the Chow Lemma twice to  $Z_N = c_{N-1}^2 \|\eta_{c_{N-1}}\|_\star^2$  and

$$a_N = \frac{\gamma_N^2}{\beta_{N-1} c_N^2} \quad (\text{resp. } Z_N = c_N \|\eta_{c_N}\|_\star \text{ and } a_N = \frac{\gamma_N^2}{\beta_{N-1} c_N}).$$

Remark that in that case, we apply the Chow Lemma to constant random variables  $(a_n)_{n \in \mathbb{N}}$ . With our hypothesis on the sequences  $(\gamma_n)_{n \in \mathbb{N}}$ ,  $(c_n)_{n \in \mathbb{N}}$  and  $(\beta_n)_{n \in \mathbb{N}}$ , the two considered sequences  $(a_n)_{n \in \mathbb{N}}$  are summable.

Theorem 4.1 is proved. Indeed, we have proved that

$$I(S^N) \xrightarrow{\text{a.s.}} I(x^*).$$

Now, one gets  $S_N \xrightarrow{\text{a.s.}} x^*$  because  $S_N$  belongs to the compact set  $C$ . If  $s^*$  is an accumulation point then because  $I(S_N) - I(x^*)$  goes to zero,  $s^*$  is also a minimum of  $I$ . So that  $s^* = x^*$  by uniqueness of the minimum.  $\square$

*Remark 8.* If some higher order moment assumption holds we could obtain an estimation of the rate of convergence, by using Markov inequality and the Chow Lemma.

## 5. APPLICATION TO RISK INDICATORS AND SIMULATIONS

We go back to our initial problem: we aim at estimating the optimal allocation of the function  $I$  (defined by (1.1)) on the simplex

$$\mathcal{U}_u = \{(v_1, \dots, v_d) \in \mathbb{R}_+^d, v_1 + \dots + v_d = u\}.$$

In order to run our mirror algorithm, we have to find an auxiliary function  $V$ , whose Legendre-Fenchel transform is computable. A natural strongly convex function on  $\mathcal{U}_u$  is the entropy function :

$$\delta_u(x) = \sum_{k=1}^d \frac{x_k}{u} \ln \left( \frac{x_k}{u} \right),$$

it is  $\frac{1}{u^2}$ -convex and admits a unique minimum  $x_0 = (\frac{u}{d}, \dots, \frac{u}{d})$ . As in [27], we consider the function

$$\begin{aligned} V(x) &= \delta_u(x) - \delta_u(x_0) - \langle \nabla \delta_u(x_0), x - x_0 \rangle \\ &= \ln d + \sum_{k=1}^d \frac{x_k}{u} \ln \left( \frac{x_k}{u} \right). \end{aligned}$$

This function is also  $\frac{1}{u^2}$ -convex and its Legendre-Fenchel transform is easily computable :

$$\nabla W_\beta(\xi) = \beta \ln \left( \frac{1}{d} \sum_{k=1}^d \exp[\xi_k \frac{u}{\beta}] \right).$$

In Proposition 2.1 above, we have proved that our risk indicator  $I$  is convex and admits a unique minimum provided that  $g_k(x) > 0$  for  $x < 0$  and that the joint distribution of  $(Y_p^k, S_p)$  contains  $] -u, +\infty[ \times ] -\infty, u[$  in its support.

The present section is devoted to simulations, whose aim is more to discuss the performances of our algorithm on simple test cases than to investigate more complex models. A more complete study on realistic models (heavy tailed distributions, temporal and spatial dependencies) will be the purpose of a future work, which should be written as a useful tool for risk managers. Here we focus on what happens when dealing with Gaussian distributions, and also when handling the so-called common shocks model. In all the examples below, the  $g_k$ 's are chosen equal to  $-\text{Id}$  :  $g_k(x) = -x$  for all  $k = 1, \dots, d$ . Moreover the cumulative gain is considered only on one period. In that case,  $n = 1$  thus  $Y = X \in \mathbb{R}^d$ .

In Section 5.1, the components of  $X$  are independent. We first consider that they also are identically distributed (see Subsection 5.1.1), then one knows that the minimum is achieved for  $u^* = (\frac{u}{d}, \dots, \frac{u}{d})$ . We can therefore study the performances of our algorithm for various tuning of the parameters,  $(\gamma_n)_{n \in \mathbb{N}}$  and  $(c_n)_{n \in \mathbb{N}}$ , on this simple test case. The sequence  $(\beta_n)_{n \in \mathbb{N}}$  is fixed equal to 1 in the following. Various choices for  $(\gamma_n)_{n \in \mathbb{N}}$  and  $(c_n)_{n \in \mathbb{N}}$  are tested, provided that they satisfy conditions (i)-(iv) of Theorem 4.1. It is well known that for standard algorithms such as Robbins-Monro or Kiefer-Wolfowitz it is hard to calibrate these sequences to get the convergence of the algorithm in practice. In the following, the sequences  $(\gamma_n)_{n \in \mathbb{N}}$  and  $(c_n)_{n \in \mathbb{N}}$  are chosen as follows :

- $\gamma_n = \frac{1}{(n+1)^\alpha}$ ,
- $c_n = \frac{1}{(n+1)^\delta}$ , with  $0 < \delta < \alpha - 1/2 < 1/2$ .

On our test cases, it seems that choosing  $\alpha$  too close to one can deteriorate the mean squared error, but the overall feeling is that our algorithm is not too sensitive to the parameters (see Subsection 5.1.1). Thus from Subsection 5.1.2, we fixed  $\alpha = \frac{3}{4} + \frac{1}{10}$  and  $\delta = \frac{1}{4}$ . In Subsection 5.1.2 we still consider independent normal coordinates for  $X$  but the marginals may have different means or variances. In Section 5.2, but some dependencies on the coordinates of  $X$  are allowed.



In section 5.3, some higher dimensional problems are considered (dimensions 10 and 40).

Also, we have chosen to perform the simulations with the initial capital  $u = d$  (except when another value for  $u$  is specified). The initialization of the algorithm ( $\chi_0$ ) is done at random uniformly in the simplex  $\mathcal{U}_u$ .

**5.1. Independent models.** The aim of this section is to provide some benchmark for the use of the algorithm. So we consider :  $d = 2$ ,  $n = 1$ , the coordinates  $X^1$  and  $X^2$  are independent. At first, we assume that  $X^1$  and  $X^2$  have the same normal distribution, then we consider different normal distributions.

**5.1.1. Same normal distribution.** We have chosen an independent normal distribution with mean equal to 0.3 and standard deviation equal to 1. For  $N = 10000$  independent simulations we obtain :  $S_N = (0.996; 1.004)$  and the following graphs. Remark that, as expected, it seems that the minimum of our risk indicator is reached for  $u_1^* = u_2^* = \frac{u}{2} = 1$ . In order to illustrate the convergence, we have plotted for each of the two coordinates the trajectory  $\chi_i$  (black line) and  $S_i$  (gray line); we have plotted the whole trajectories ( $i = 1, \dots, 10000$ ) and a zoom on the end of the trajectories ( $i = 2000, \dots, 10000$ ). We see on these graphs the importance of the averaging step which provides a solution which is much more smooth.

FIGURE 1. CONVERGENCE OF THE ALGORITHM, FIRST COORDINATE

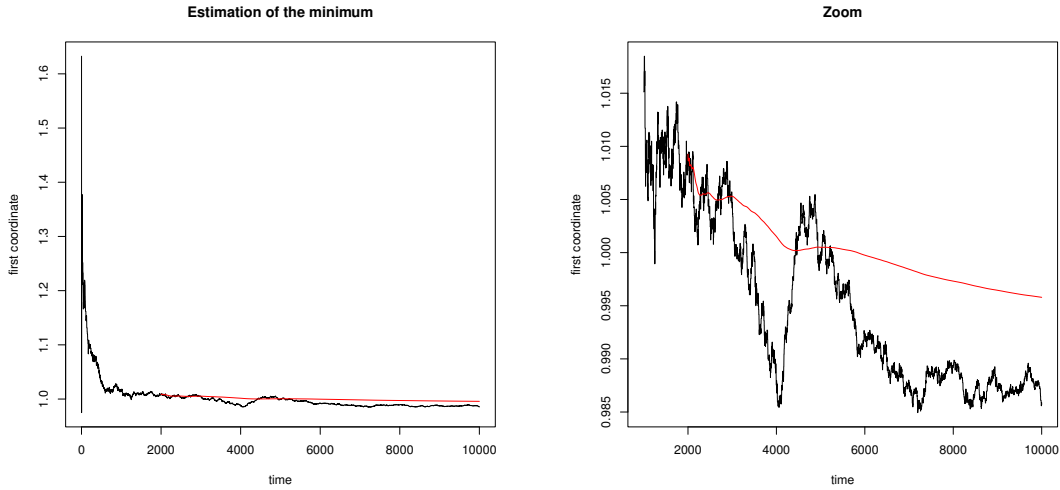
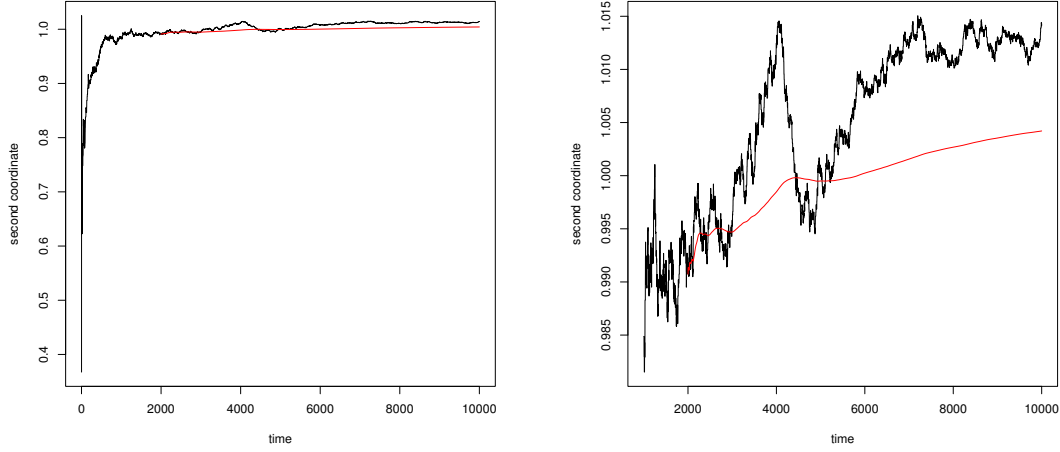


FIGURE 2. CONVERGENCE OF THE ALGORITHM, SECOND COORDINATE  
 Estimation of the minimum                      Zoom



We conclude this subsection by a study of  $k = 30$  simulations of length  $N = 1000$ , with various sequences  $(\gamma_n)_{n \in \mathbb{N}}$  and  $(c_n)_{n \in \mathbb{N}}$ . These simulations allow to see the influence of these sequences on the results of the algorithms. The performance is measured with the mean squared error :

$$mse = \frac{1}{k} \sum_{j=1}^k \|u^* - \hat{u}^j\|^2,$$

where  $\hat{u}^j$  is the estimated value of the vector  $u^*$  on the  $j^{\text{th}}$  simulation.

The table below gives for each of the two coordinates, the mean of the estimation  $(\hat{u}_1, \hat{u}_2)$  of the minimum  $(u_1^*, u_2^*)$ , the standard error and the mean squared error  $mse$  for sequences  $(\gamma_n)_{n \in \mathbb{N}}$  and  $(c_n)_{n \in \mathbb{N}}$  :

$\gamma_n = \frac{1}{(n+1)^\alpha}$  and  $c_n = \frac{1}{(n+1)^\delta}$  for various values of  $\alpha$  and  $\delta$  satisfying the hypothesis of Theorem 4.1. The parameter  $\delta$  is used for finite differences and it seems that it is better to choose it not too close to 1, as it deteriorates the mean squared error. The parameter  $\alpha$  (used to define the descent step  $(\gamma_n)_{n \in \mathbb{N}}$ ) does not seem to have a great influence on our test case.

	$\alpha = 1$ $\delta = \frac{1}{2}$	$\alpha = 1$ $\delta = \frac{1}{4}$	$\alpha = \frac{3}{4} + \frac{1}{10}$ $\delta = \frac{1}{2}$	$\alpha = \frac{3}{4} + \frac{1}{10}$ $\delta = \frac{1}{4}$	$\alpha = \frac{3}{4}$ $\delta = \frac{1}{2}$	$\alpha = \frac{3}{4}$ $\delta = \frac{1}{4}$
1 <sup>st</sup> coord. mean	1.02	1.01	1.01	1.02	1.01	1.01
1 <sup>st</sup> standard error	0.084	0.071	0.04	0.04	0.03	0.03
2 <sup>nd</sup> coord. mean	0.98	0.99	0.99	0.98	0.99	0.99
2 <sup>nd</sup> standard error	0.084	0.071	0.04	0.04	0.03	0.03
mse	0.015	0.01	0.003	0.003	0.002	0.002

However, even with a naive fitting of the parameters we observe the convergence of the algorithm in practice.

**5.1.2. Different normal distributions.** For this subsection, we have chosen to present the results of 50 simulations of length  $N = 1000$ , with two independent normal distributions. First of all, we consider for the first coordinate a normal distribution  $\mathcal{N}(0.3, 1)$  and for the second one a normal distribution  $\mathcal{N}(0.8, 1)$ .

The table below gives for each of the two coordinates, the mean of the estimation  $(\widehat{u}_1, \widehat{u}_2)$  of the minimum  $(u_1^*, u_2^*)$ .

	first coord.	second coord.
mean	1.226	0.774
standard error	0.051	0.051

We observe that there is a significant difference between the two coordinates. This is intuitive that in order to minimize the risk indicator, one should affect more initial capital to the branch whose expected income is the less (which is the more risky branch in that case).

We achieve this section with independent normal distributions with different variance (and same mean). As above, we have performed 50 simulations of length  $N = 1000$ , with two independent normal distributions. For the first coordinate we consider  $\mathcal{N}(0.3, 1)$  and for the second one  $\mathcal{N}(0.3, 4)$ .

The table below gives for each of the two coordinates, the mean of the estimation  $(\widehat{u}_1, \widehat{u}_2)$  of the minimum  $(u_1^*, u_2^*)$  and the standard error.

	first coord.	second coord.
mean	0.787	1.213
standard error	0.067	0.067

In that case also, we observe that there is a significant difference between the two coordinates. Here, the more risky branch is the second one (which has a higher variance).

**5.2. Models with spatial dependencies.** We propose two simple models of spatial dependence. First of all we consider random vectors in  $\mathbb{R}^3$  with common shocks. A second example of spatial dependence is given by Gaussian vectors.

**5.2.1. Model with common shocks.** Models with common shocks are very popular in insurance because they allow to model dependencies due to some common factor. We consider a dimension three vector:

$$X_1 = I \times Z_1 + (1 - I) \times W, \quad X_2 = I \times Z_2 + (1 - I) \times W, \quad X_3 = Z_3$$

with  $Z_1, Z_2, Z_3$  independent normal distributions  $\mathcal{N}(0.3, 1)$ ,  $W = 0.3 + T$ ,  $T$  a Student distribution with 5 degrees of freedom and  $I$  a Bernoulli distribution with parameter  $\frac{1}{5}$ .

The  $Z$ 's,  $W$ ,  $I$  are independent.

We have simulated 30 times 1000 independent periods of length  $n = 1$  with  $u = 2$ .

	first coord.	second coord.	third coord.
mean	0.61	0.61	0.78
standard error	0.024	0.033	0.042

The first two coordinates are positively correlated. If one fails then the probability that the other one fails is more important. Putting more capital for the entity constituted with these first two branches should reduce the insolvency risk of the branches.

5.2.2. *Gaussian vectors.* We conclude this paragraph on vectorial dependence with an example of a Gaussian vector in dimension 2 and an example of a Gaussian vector in dimension 3. First, we consider the Gaussian vector with covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$$

and expectation  $m = (0.3, 0.3)$ . As above, we have performed 50 simulations of length  $N = 1000$ , of the Gaussian vector  $X$ .

The table below gives for each of the two coordinates, the mean of the estimation  $(\widehat{u}_1, \widehat{u}_2)$  of the minimum  $(u_1^*, u_2^*)$  and the standard error.

	first coord.	second coord.
mean	0.999	1.001
standard error	0.034	0.034

It seems that the result is the same as in the non correlated case but with a smaller standard deviation.

We have also performed the simulations for a Gaussian vector with expectation  $m = (0.3, 0.8)$  and the covariance matrix  $\Sigma$ . The table below gives for each of the two coordinates, the mean of the estimation  $(\widehat{u}_1, \widehat{u}_2)$  of the minimum  $(u_1^*, u_2^*)$  and the standard error.

	first coord.	second coord.
mean	1.21	0.79
standard error	0.034	0.034

As in the above example where the marginal distributions were the same, it seems that the result is similar to the non correlated case but with a smaller standard error.

We have also performed simulations for a Gaussian vector in  $\mathbb{R}^3$  with covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.9 \\ 0 & 0.9 & 1 \end{pmatrix}$$

and expectation  $m = (0.3, 0.3, 0.3)$ . As above, we have performed 50 simulations of length  $N = 1000$ , of the Gaussian vector  $X$ .

The table below gives for each of the three coordinates, the mean of the estimation  $(\widehat{u}_1, \widehat{u}_2, \widehat{u}_3)$  of the minimum  $(u_1^*, u_2^*, u_3^*)$  and the standard error.

	first coord.	second coord.	third coord.
mean	0.785	0.604	0.612
standard error	0.045	0.03	0.028

As in the example with common shocks, the strategy is to put more capital on the entity constituted with the positively correlated branches.

**5.3. Simulations in higher dimension.** We have performed simulations in higher dimension, still considering one period for the cumulative gain ( $n = 1$ ). First for Gaussian random vectors in dimension 10 then in dimension 40, in order to see if our algorithm gives good results in high dimension. In order to measure the performance of the algorithm, we have performed  $k = 30$  simulations of length  $N = 1000$ , with independent normal Gaussian vectors of  $\mathbb{R}^{10}$  then of  $\mathbb{R}^{40}$ , each coordinate following a normal  $\mathcal{N}(0.3, 1)$ . We have chosen for the sequences  $\gamma_n = \frac{1}{(n+1)}$  and  $c_n = \frac{1}{(n+1)^{\frac{1}{4}}}$ .

**5.3.1. Results in dimension 10.** The table below contains for each coordinate, the mean - over the 30 simulations - of the estimated value of  $\hat{u}_i$  and the standard-error. Then we have computed the mse. Remark that as before, since the coordinates of the vector are i.i.d. then the minimum of the indicator will be realized for  $u_i^* = \frac{u}{10} = 1$  for  $i = 1, \dots, 10$ .

coordinate	mean	standard error
1	0.997	0.043
2	1.004	0.044
3	1.002	0.042
4	0.996	0.042
5	0.987	0.044
6	1.004	0.05
7	0.998	0.044
8	1.002	0.044
9	1.016	0.045
10	0.993	0.048

The mean squared error is 0.02. Even if the mean squared is greater (around 5 times greater as in dimension 2), it seems that the algorithm is still performant in dimension 10.

We have also considered in dimension 10, a Gaussian vector with mean 0.3 for all the coordinates and with the covariance matrix:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \end{pmatrix}$$

The table below contains for each coordinate, the mean - over the 30 simulations - of the estimated value of  $\hat{u}_i$  and the standard-error.

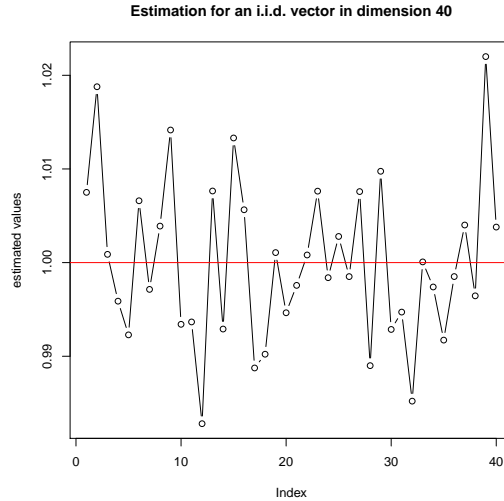
coordinate	mean	standard error
1	1.192	0.043
2	1.19	0.035
3	1.188	0.034
4	1.189	0.034
5	1.189	0.034
6	0.81	0.036
7	0.811	0.035
8	0.81	0.035
9	0.81	0.036
10	0.811	0.034

Clearly, the algorithm recognizes two blocks in the random vector and puts more capital on the more risky (with higher variance) one.

**5.3.2. Results in dimension 40.** We conclude this simulation session by a simple example in dimension 40. We have performed  $k = 30$  simulations of length  $N = 1000$ , with independent normal Gaussian vectors each coordinate following a normal distribution  $\mathcal{N}(0.3, 1)$ . We have chosen for the sequences  $\gamma_n = \frac{1}{(n+1)}$  and  $c_n = \frac{1}{(n+1)^{\frac{1}{4}}}$ . The mean squared error is 0.09.

The picture below shows the mean (over the 30 simulations) of the estimated values  $\hat{u}_1, \dots, \hat{u}_{40}$  of  $u_1^* = \dots = u_{40}^* = \frac{u}{40} = 1$ .

FIGURE 3. ESTIMATED VALUES OF  $u_1^*, \dots, u_{40}^*$ .



All these results show that our algorithm is tractable even for relative high dimensions. Of course the mean squared error increases with the dimension but even in dimension 40 we obtain a quite good precision for our estimator (it seem that the  $mse$  increases linearly with the dimension). Note that for all these simulations, we did not encounter any difficulty for fitting our various parameters.

More realistic models, including temporal dependencies and heavy tailed

distributions, have to be investigated from a theoretical and simulation point of view, in order to propose a useful tool for risk managers. This will be the aim of a forthcoming work.

**Acknowledgments.** We are very grateful to Anatoli Juditsky for his valuable advices. Moreover, we would like to thank the referees for their careful reading of our manuscript and their comments.

## REFERENCES

- [1] P. Artzner, F. Delbaen, J-M. Eber, D. Heath *Coherent Measures of Risk.*, Mathematical Finance 9 (3), 203-228, (1999).
- [2] N. Bartoli, P. Del Moral *Simulation & algorithmes stochastiques*. Cépaduès-éditions (2001).
- [3] R. Biard, S. Loisel, C. Macci and N. Veraverbeke *Asymptotic behavior of the finite-time expected time-integrated negative part of some risk processes and optimal reserve allocation*. Journal of Mathematical Analysis and Applications, to appear (2010).
- [4] J. Cai, H. Li *Dependence properties and bounds for ruin probabilities in multivariate compound risk models*. Journal of Multivariate Analysis, **98**, 757-773 (2007).
- [5] E. Di Bernadino, T. Laloë, V. Maume-Deschamps, C. Prieur *Plug-in estimation of level sets in a non-compact setting with applications in multivariate risk theory*. Submitted for publication, available at <http://hal.archives-ouvertes.fr/hal-00580624/fr/>.
- [6] F. Dufresne, H.U. Gerber, *The surpluses immediately before and at ruin, and the amount of the claim causing ruin*. Insurance Math. Econom. 7-3, 193-199 (1998).
- [7] M. Duflo, *Random Iterative Models*. Springer (1997).
- [8] H.U. Gerber, *An introduction to Mathematical Risk Theory*. Huebner Foundation monograph n°8, Irwin Homewood IL (1979).
- [9] S. Ghdimi & G. Lan, *Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization*, Manuscript, Georgia Institute of Technology, (2010).
- [10] H. Hult, F. Lindskog, *Heavy-tailed insurance portfolios: buffer capital and ruin probabilities*. Technical Report 1441, School of ORIE, Cornell University (2006).
- [11] A. Juditsky, A. Nazin, A.B. Tsybakov, N. Vayatis, *Recursive aggregation of estimators via the mirror descent algorithm with average*. Problems of Information Transmission, 41:n.4, (2005)
- [12] A. Juditsky, P. Rigollet, A.B. Tsybakov, *Learning by mirror averaging*. Annals of Statistics, 36:2183-2206, (2008).
- [13] J. Kiefer, J. Wolfowitz *Stochastic estimation of the maximum of a regression function*. Ann. Math. Statistics 23, (1952). 462-466.
- [14] HJ Kushner, G Yin, *Stochastic approximation and recursive algorithms and applications.*, Applications of Mathematics Stochastic modelling and applied probability (2003).
- [15] G. Lan, *Efficient methods for stochastic composite optimization*. Manuscript, Georgia Institute of Technology, (2008),
- [16] G. Lan *An optimal method for stochastic composite optimization*. to appear in Mathematical Programming (2010).
- [17] J. Lelong, *Almost sure convergence of randomly truncated stochastic algorithms under verifiable conditions*. Statistics & Probability Letters, **78**, 16, 2632-2636 (2008).
- [18] J. Neveu *Martingales à temps discret*. Masson et Cie, éditeurs, Paris, (1972).
- [19] Nemirovski & Yudin, *Problem Complexity and Method Efficiency in Optimization*. John Wiley and Sons, (1983).
- [20] A. Nemirovski, A. Juditsky, G. Lan, A. Shapiro, *Robust stochastic approximation to stochastic programming*, SIAM, Journal on Optimization, 19:1574-1609, (2009).
- [21] Nesterov, *Primal-dual subgradient methods for convex problems.*, Mathematical Programming, 120:221-259, (2006)

- [22] M. Pelletier, *Asymptotic almost sure efficiency of averaged stochastic algorithms*. SIAM J. Control Optim, 39(1):49–72 (2000).
- [23] P. Picard, *On some measures of the severity of ruin in the classical Poisson model*. Insurance Math. Econom. 14-2, 107-115, (1994).
- [24] Polyak & Juditsky *Acceleration of stochastic approximation by averaging* SIAM, J. Control and Optimization, 30:838-855, (1992).
- [25] Robbins & Monro *A stochastic approximation method.*, Annals of Mathematical Statistics, 22:400-407, (1951)
- [26] S. Loisel, *Differentiation of some functionals of risk processes and optimal reserve allocation*. Journal of Applied Probability, **42**, 2, 379-392 (2005).
- [27] C. Tauvel, *Optimisation stochastique à grande échelle*, PhD Thesis, (2008).

UNIVERSITÉ DE BOURGOGNE, IMB UMR 5584  
*E-mail address:* `peggy.cenac@u-bourgogne.fr`

UNIVERSITÉ DE LYON, UNIVERSITÉ LYON 1, LABORATOIRE SAF EA 2429  
*E-mail address:* `veronique.maume@univ-lyon1.fr`

UNIVERSITÉ JOSEPH FOURIER, TOUR IRMA, MOISE-LJK  
*E-mail address:* `clementine.prieur@imag.fr`